

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: McKee, Marian L.  
O'Brien, Alison D.  
Wachtel, Marian R.
- (ii) TITLE OF INVENTION: Histidine-Tagged Intimin And Methods Of  
Using Intimin To Stimulate An Immune Response And As An  
Antigen Carrier With Targeting Capability.
- (iii) NUMBER OF SEQUENCES: 25
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
Dunner, L.L.P.
  - (B) STREET: 1300 I Street, N.W., Suite 700
  - (C) CITY: Washington
  - (D) STATE: D.C.
  - (E) COUNTRY: USA
  - (F) ZIP: 20005-3315
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/837,459
  - (B) FILING DATE: 18-APR-1997
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Fordis, Jean B.
  - (B) REGISTRATION NUMBER: 32,984
  - (C) REFERENCE/DOCKET NUMBER: 04995.0023-00000
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (202) 408-4000
  - (B) TELEFAX: (202) 408-4400

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGTTGTTAAG TCAATGGAAA C

21

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TCTAGAGAGA AAACGTGAAT GTTGTCTCT

29

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTACGGATCC ATGATGGTTT TCCAGCCAAT CAGTGAG

37

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GTACGGTACC TTATATTGAC AGCGCACAGA GCGGG

35

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTACGGATCC ATATGTGGAA TGTTTCATGGC TGGGG

35

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GTACGGATCC GAATTCATTT GCAAATGGTG

30

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTACGGTACC TGATCAATGA AGACGTTATA G

31

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GTACGGATCC TGATCAGGAT TTTTCTGGTG

30

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 30 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTACGGTACC TGATCAAAAA ATATAACCGC

30

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 34 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTACGGATCC TGATCAAACC AAGGCCAGCA TTAC

34

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 32 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTACGGTACC TTATTCTACA CAAACCGCAT AG

32

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GTACGGATCC ACTGAAAGCA AGCGGTGGTG ATG

33

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 31 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GTACGGATCC TTCATGGTAT TCAGAAAATA C

31

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTACGGATCC GACTGTCGAT GCATCAGGGA AAG

33

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTACGGATCC GAATGGTAAA GGCAGTGTCTG

30

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GTACGGTACC TCCAGAACGC TGCTCACTAG

30

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTACGGTACC TTATTCTACA GAAACGCGAT AG

32

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATAACATGAG TACTCATGGT TG

22

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 934 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met	Ile	Thr	His	Gly	Cys	Tyr	Thr	Arg	Thr	Arg	His	Lys	His	Lys	Leu
1				5					10					15	

Lys	Lys	Thr	Leu	Ile	Met	Leu	Ser	Ala	Gly	Leu	Gly	Leu	Phe	Phe	Tyr
			20						25					30	

Val	Asn	Gln	Asn	Ser	Phe	Ala	Asn	Gly	Glu	Asn	Tyr	Phe	Lys	Leu	Gly
			35					40					45		

Ser	Asp	Ser	Lys	Leu	Leu	Thr	His	Asp	Ser	Tyr	Gln	Asn	Arg	Leu	Phe
			50				55					60			

Tyr	Thr	Leu	Lys	Thr	Gly	Glu	Thr	Val	Ala	Asp	Leu	Ser	Lys	Ser	Gln
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

65		70		75		80									
Asp	Ile	Asn	Leu	Ser	Thr	Ile	Trp	Ser	Leu	Asn	Lys	His	Leu	Tyr	Ser
				85					90					95	
Ser	Glu	Ser	Glu	Met	Met	Lys	Ala	Ala	Pro	Gly	Gln	Gln	Ile	Ile	Leu
			100					105					110		
Pro	Leu	Lys	Lys	Leu	Pro	Phe	Glu	Tyr	Ser	Ala	Leu	Pro	Leu	Leu	Gly
		115					120					125			
Ser	Ala	Pro	Leu	Val	Ala	Ala	Gly	Gly	Val	Ala	Gly	His	Thr	Asn	Lys
	130					135					140				
Leu	Thr	Lys	Met	Ser	Pro	Asp	Val	Thr	Lys	Ser	Asn	Met	Thr	Asp	Asp
145					150					155					160
Lys	Ala	Leu	Asn	Tyr	Ala	Ala	Gln	Gln	Ala	Ala	Ser	Leu	Gly	Ser	Gln
			165						170					175	
Leu	Gln	Ser	Arg	Ser	Leu	Asn	Gly	Asp	Tyr	Ala	Lys	Asp	Thr	Ala	Leu
			180					185					190		
Gly	Ile	Ala	Gly	Asn	Gln	Ala	Ser	Ser	Gln	Leu	Gln	Ala	Trp	Leu	Gln
	195						200					205			
His	Tyr	Gly	Thr	Ala	Glu	Val	Asn	Leu	Gln	Ser	Gly	Asp	Asn	Phe	Asp
	210					215					220				
Gly	Ser	Ser	Leu	Asp	Phe	Leu	Leu	Pro	Phe	Tyr	Asp	Ser	Glu	Lys	Met
225					230					235					240
Leu	Ala	Phe	Gly	Gln	Val	Gly	Ala	Arg	Tyr	Ile	Asp	Ser	Arg	Phe	Thr
			245						250					255	
Ala	Asn	Leu	Gly	Ala	Gly	Gln	Arg	Phe	Phe	Leu	Pro	Ala	Asn	Met	Leu
			260					265					270		
Gly	Tyr	Asn	Val	Phe	Ile	Asp	Gln	Asp	Phe	Ser	Gly	Asp	Asn	Thr	Arg
	275						280					285			
Leu	Gly	Ile	Gly	Gly	Glu	Tyr	Trp	Arg	Asp	Tyr	Phe	Lys	Ser	Ser	Val
	290					295					300				
Asn	Gly	Tyr	Phe	Arg	Met	Arg	Arg	Trp	His	Glu	Ser	Tyr	His	Lys	Lys
305					310					315					320
Asp	Tyr	Asp	Glu	Arg	Pro	Ala	Asn	Gly	Phe	Asp	Ile	Arg	Phe	Asn	Gly
			325						330					335	
Tyr	Leu	Pro	Ser	Tyr	Pro	Ala	Leu	Gly	Ala	Lys	Leu	Ile	Tyr	Glu	Gln
			340					345					350		
Tyr	Tyr	Gly	Asp	Asn	Val	Ala	Leu	Phe	Asn	Ser	Asp	Lys	Leu	Gln	Ser



355								360								365
Asn	Pro	Gly	Ala	Ala	Thr	Val	Gly	Val	Asn	Tyr	Thr	Pro	Ile	Pro	Leu	
370							375				380					
Val	Thr	Met	Gly	Ile	Asp	Tyr	Arg	His	Gly	Thr	Gly	Asn	Glu	Asn	Asp	
385					390					395					400	
Leu	Leu	Tyr	Ser	Met	Gln	Phe	Arg	Tyr	Gln	Phe	Asp	Lys	Ser	Trp	Ser	
				405					410					415		
Gln	Gln	Ile	Glu	Pro	Gln	Tyr	Val	Asn	Glu	Leu	Arg	Thr	Leu	Ser	Gly	
			420					425					430			
Ser	Arg	Tyr	Asp	Leu	Val	Gln	Arg	Asn	Asn	Asn	Ile	Ile	Leu	Glu	Tyr	
		435					440					445				
Lys	Lys	Gln	Asp	Ile	Leu	Ser	Leu	Asn	Ile	Pro	His	Asp	Ile	Asn	Gly	
	450					455					460					
Thr	Glu	His	Ser	Thr	Gln	Lys	Ile	Gln	Leu	Ile	Val	Lys	Ser	Lys	Tyr	
465					470					475					480	
Gly	Leu	Asp	Arg	Ile	Val	Trp	Asp	Asp	Ser	Ala	Leu	Arg	Ser	Gln	Gly	
				485					490					495		
Gly	Gln	Ile	Gln	His	Ser	Gly	Ser	Gln	Ser	Ala	Gln	Asp	Tyr	Gln	Ala	
			500					505					510			
Ile	Leu	Pro	Ala	Tyr	Val	Gln	Gly	Gly	Ser	Asn	Ile	Tyr	Lys	Val	Thr	
		515					520					525				
Ala	Arg	Ala	Tyr	Asp	Arg	Asn	Gly	Asn	Ser	Ser	Asn	Asn	Val	Gln	Leu	
	530					535					540					
Thr	Ile	Thr	Val	Leu	Ser	Asn	Gly	Gln	Val	Val	Asp	Gln	Val	Gly	Val	
545					550					555					560	
Thr	Asp	Phe	Thr	Ala	Asp	Lys	Thr	Ser	Ala	Lys	Ala	Asp	Asn	Ala	Asp	
				565					570					575		
Thr	Ile	Thr	Tyr	Thr	Ala	Thr	Val	Lys	Lys	Asn	Gly	Val	Ala	Gln	Ala	
			580					585					590			
Asn	Val	Pro	Val	Ser	Phe	Asn	Ile	Val	Ser	Gly	Thr	Ala	Thr	Leu	Gly	
		595					600					605				
Ala	Asn	Ser	Ala	Lys	Thr	Asp	Ala	Asn	Gly	Lys	Ala	Thr	Val	Thr	Leu	
	610					615					620					
Lys	Ser	Ser	Thr	Pro	Gly	Gln	Val	Val	Val	Ser	Ala	Lys	Thr	Ala	Glu	
625					630					635					640	
Met	Ser	Ser	Ala	Leu	Asn	Ala	Ser	Ala	Val	Ile	Phe	Phe	Asp	Gln	Thr	

645								650				655			
Lys	Ala	Ser	Ile	Thr	Glu	Ile	Lys	Ala	Asp	Lys	Thr	Thr	Ala	Val	Ala
660								665				670			
Asn	Gly	Lys	Asp	Ala	Ile	Lys	Tyr	Thr	Val	Lys	Val	Met	Lys	Asn	Gly
675								680				685			
Gln	Pro	Val	Asn	Asn	Gln	Ser	Val	Thr	Phe	Ser	Thr	Asn	Phe	Gly	Met
690								695				700			
Phe	Asn	Gly	Lys	Ser	Gln	Thr	Gln	Ala	Thr	Thr	Gly	Asn	Asp	Gly	Arg
705								710				715			
Ala	Thr	Ile	Thr	Leu	Thr	Ser	Ser	Ser	Ala	Gly	Lys	Ala	Thr	Val	Ser
725								730				735			
Ala	Thr	Val	Ser	Asp	Gly	Ala	Glu	Val	Lys	Ala	Thr	Glu	Val	Thr	Phe
740								745				750			
Phe	Asp	Glu	Leu	Lys	Ile	Asp	Asn	Lys	Val	Asp	Ile	Ile	Gly	Asn	Asn
755								760				765			
Val	Arg	Gly	Glu	Leu	Pro	Asn	Ile	Trp	Leu	Gln	Tyr	Gly	Gln	Phe	Lys
770								775				780			
Leu	Lys	Ala	Ser	Gly	Gly	Asp	Gly	Thr	Tyr	Ser	Trp	Tyr	Ser	Glu	Asn
785								790				795			
Thr	Ser	Ile	Ala	Thr	Val	Asp	Ala	Ser	Gly	Lys	Val	Thr	Leu	Asn	Gly
805								810				815			
Lys	Gly	Ser	Val	Val	Ile	Lys	Ala	Thr	Ser	Gly	Asp	Lys	Gln	Thr	Val
820								825				830			
Ser	Tyr	Thr	Ile	Lys	Ala	Pro	Ser	Tyr	Met	Ile	Lys	Val	Asp	Lys	Gln
835								840				845			
Ala	Tyr	Tyr	Ala	Asp	Ala	Met	Ser	Ile	Cys	Lys	Asn	Leu	Leu	Pro	Ser
850								855				860			
Thr	Gln	Thr	Val	Leu	Ser	Asp	Ile	Tyr	Asp	Ser	Trp	Gly	Ala	Ala	Asn
865								870				875			
Lys	Tyr	Ser	His	Tyr	Ser	Ser	Met	Asn	Ser	Ile	Thr	Ala	Trp	Ile	Lys
885								890				895			
Gln	Thr	Ser	Ser	Glu	Gln	Arg	Ser	Gly	Val	Ser	Ser	Thr	Tyr	Asn	Leu
900								905				910			
Ile	Thr	Gln	Asn	Pro	Leu	Pro	Gly	Val	Asn	Val	Asn	Thr	Pro	Asn	Val
915								920				925			
Tyr Ala Val Cys Val Glu															

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3131 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TCGAGAATGA AATAGAAGTC GTTGTTAAGT CAATGGAAAA CCTGTATTTG GTATTACATA	60
ATCAGGGAAT AACATTAGAA AACGAACATA TGAAAATAGA GGAAATCAGT TCAAGCGACA	120
ATAAACATTA TTACGCCGGA AGATAAAATC CGATCTATTA ATATAATTTA TTTCTCATTC	180
TAACTCATTG TGGTGGAGCC ATAACATGAT TACTCATGGT TGTTATACCC GGACCCGGCA	240
CAAGCATAAG CTAAAAAAAA CATTGATTAT GCTTAGTGCT GGTTTAGGAT TGTTTTTTTA	300
TGTTAATCAG AATTCATTTG CAAATGGTGA AAATTATTTT AAATTGGGTT CGGATTCAAA	360
ACTGTTAACT CATGATAGCT ATCAGAATCG CCTTTTTTAT ACGTTGAAAA CTGGTGAAAC	420
TGTTGCCGAT CTTTCTAAAT CGCAAGATAT TAATTTATCG ACGATTTGGT CGTTGAATAA	480
GCATTTATAC AGTTCTGAAA GCGAAATGAT GAAGGCCGCG CCTGGTCAGC AGATCATTTT	540
GCCACTCAAA AAACCTCCCT TTGAATACAG TGCACTACCA CTTTTAGGTT CGGCACCTCT	600
TGTTGCTGCA GGTGGTGTG CTGGTCACAC GAATAAACTG ACTAAAATGT CCCC GGACGT	660
GACCAAAAGC AACATGACCG ATGACAAGGC ATTAAATTAT GCGGCACAAC AGGCGGCGAG	720
TCTCGGTAGC CAGCTTCAGT CGCGATCTCT GAACGGCGAT TACGCGAAAG ATACCGCTCT	780
TGGTATCGCT GGTAACCAGG CTTTCGTCACA GTTGCAGGCC TGGTTACAAC ATTATGGAAC	840
GGCAGAGGTT AATCTGCAGA GTGGTAATAA CTTTGACGGT AGTTCACTGG ACTTCTTATT	900
ACCGTTCTAT GATTCCGAAA AAATGCTGGC ATTTGGTCAG GTCGGAGCGC GTTACATTGA	960
CTCCCGCTTT ACGGCAAATT TAGGTGCGGG TCAGCGTTTT TTCCTTCCTG CAAACATGTT	1020
GGGCTATAAC GTCTTCATTG ATCAGGATTT TTCTGGTGAT AATACCCGTT TAGGTATTGG	1080
TGGCGAATAC TGGCGAGACT ATTTCAAAAAG TAGCGTTAAC GGCTATTTCC GCATGAGCGG	1140

CTGGCATGAG	TCATACAATA	AGAAAGACTA	TGATGAGCGC	CCAGCAAATG	GCTTCGATAT	1200
CCGTTTTAAT	GGCTATCTAC	CGTCATATCC	GGCATTAGGC	GCCAAGCTGA	TATATGAGCA	1260
GTATTATGGT	GATAATGTTG	CTTTGTTTAA	TTCTGATAAG	CTGCAGTCGA	ATCCTGGTGC	1320
GGCGACCGTT	GGTGTAAGCT	ATACTCCGAT	TCCTCTGGTG	ACGATGGGGA	TCGATTACCG	1380
TCATGGTACG	GGTAATGAAA	ATGATCTCCT	TTACTCAATG	CAGTTCCGTT	ATCAGTTTGA	1440
TAAATCGTGG	TCTCAGCAAA	TTGAACCACA	GTATGTTAAC	GAGTTAAGAA	CATTATCAGG	1500
CAGCCGTTAC	GATCTGGTTC	AGCGTAATAA	CAATATTATT	CTGGAGTACA	AGAAGCAGGA	1560
TATTCTTTCT	CTGAATATTC	CGCATGATAT	TAATGGTACT	GAACACAGTA	CGCAGAAGAT	1620
TCAGTTGATC	GTTAAGAGCA	AATACGGTCT	GGATCGTATC	GTCTGGGATG	ATAGTGCATT	1680
ACGCAGTCAG	GGCGGTGAGA	TTCAGCATAG	CGGAAGCCAA	AGCGCACAAAG	ACTACCAGGC	1740
TATTTTGCCT	GCTTATGTGC	AAGGTGGCAG	CAATATTTAT	AAAGTGACGG	CTCGCGCCTA	1800
TGACCGTAAT	GGCAATAGCT	CTAACAATGT	ACAGCTTACT	ATTACCGTTC	TGTCGAATGG	1860
TCAAGTTGTC	GACCAGGTTG	GGGTAACGGA	CTTTACGGCG	GATAAGACTT	CGGCTAAAGC	1920
GGATAACGCC	GATACCATTA	CTTATACCGC	GACGGTGAAA	AAGAATGGGG	TAGCTCAGGC	1980
TAATGTCCCT	GTTTCATTTA	ATATTGTTTC	AGGAACTGCA	ACTCTTGGGG	CAAATAGTGC	2040
CAAAACGGAT	GCTAACGGTA	AGGCAACCGT	AACGTTGAAG	TCGAGTACGC	CAGGACAGGT	2100
CGTCGTGTCT	GCTAAAACCG	CGGAGATGAC	TTCAGCACTT	AATGCCAGTG	CGGTTATATT	2160
TTTTGATCAA	ACCAAGGCCA	GCATTACTGA	GATTAAGGCT	GATAAGACAA	CTGCAGTAGC	2220
AAATGGTAAG	GATGCTATTA	AATATACTGT	AAAAGTTATG	AAAAACGGTC	AGCCAGTTAA	2280
TAATCAATCC	GTTACATTCT	CAACAACTT	TGGGATGTTT	AACGGTAAGT	CTCAAACGCA	2340
AGCAACCACG	GGAAATGATG	GTCGTGCGAC	GATAACACTA	ACTTCCAGTT	CCGCCGGTAA	2400
AGCGACTGTT	AGTGCGACAG	TCAGTGATGG	GGCTGAGGTT	AAAGCGACTG	AGGTCACTTT	2460
TTTTGATGAA	CTGAAAATTG	ACAACAAGGT	TGATATTATT	GGTAACAATG	TCAAGAGGTC	2520
GATGTTGCCT	AATATTTGGC	TGCAATATGG	TCAGTTTAAA	CTGAAAGCAA	GCGGTGGTGA	2580
TGGTACATAT	TCATGGTATT	CAGAAAATAC	CAGTATCGCG	ACTGTTCGATG	CATCAGGGAA	2640
AGTCACTTTG	AATGGTAAAG	GCAGTGTCGT	AATTAAAGCC	ACATCTGGTG	ATAAGCAAAC	2700
AGTAAGTTAC	ACTATAAAAG	CACCGTCGTA	TATGATAAAA	GTGGATAAGC	AAGCCTATTA	2760

TGCTGATGCT ATGTCCATTT GCAAAAATTT ATTACCATCC ACACAGACGG TATTGTCAGA	2820
TATTTATGAC TCATGGGGGG CTGCAAATAA ATATAGCCAT TATAGTTCTA TGAACCTCAAT	2880
AACTGCTTGG ATTAAACAGA CATCTAGTGA GCAGCGTTCT GGAGTATCAA GCACTTATAA	2940
CCTAATAACA CAAAACCCCTC TTCCTGGGGT TAATGTTAAT ACTCCAAATG TCTATGCGGT	3000
TTGTGTAGAA TAATTCCATA ACCACCCCGG CTAAAATATG TATTGTTTTA GTCGGGGCAT	3060
AATTATTTCT TCTTAAGAAA TAACCCTCTT ATAATCAAAT CTACTACTGG TCTTTTTATC	3120
TGCTTAATAG G	3131

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3106 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGAAAGATAA ATCCGATCTA TTAATATAAT TTATTTCTCA TTCTAACTCA TTGTGGTGGGA	60
GCCATAACAT GAGTACTCAT GGTGTGTATA CCCGGACCCG GCACAAGCAT AAGCTAAAAA	120
AAACATTGAT TATGCTTAGT GCTGGTTTAG GATTGTTTTT TTATGTTAAT CAGAATTCAT	180
TTGCAAATGG TGAAAATTAT TTAAATTGG GTTCGGATTC AAAACTGTTA ACTCATGATA	240
GCTATCAGAA TCGCCTTTTT TATACGTTGA AAACTGGTGA AACTGTTGCC GATCTTTCTA	300
AATCGCAAGA TATTAATTTA TCGACGATTT GGTCGTTGAA TAAGCATTTA TACAGTTCTG	360
AAAGCGAAAT GATGAAGGCC GCGCCTGGTC AGCAGATCAT TTTGCCACTC AAAAACTTC	420
CCTTTGAATA CAGTGCACTA CCACTTTTAG GTTCGGCACC TCTTGTTGCT GCAGGTGGTG	480
TTGCTGGTCA CACGAATAAA CTGACTAAAA TGTCCCCGGA CGTGACCAA AGCAACATGA	540
CCGATGACAA GGCATTAAAT TATGCGGCAC AACAGGCGGC GAGTCTCGGT AGCCAGCTTC	600
AGTCGCGATC TCTGAACGGC GATTACGCGA AAGATACCGC TCTTGGTATC GCTGGTAACC	660
AGGCTTCGTC ACAGTTGCAG GCCTGGTTAC AACATTATGG AACGGCAGAG GTTAATCTGC	720
AGAGTGGTGA TAACTTTGAC GGTAGTTCAC TGGACTTCTT ATTACCGTTC TATGATTCCG	780

AAAAAATGCT	GGCATTGTTGGT	CAGGTCGGAG	CGCGTTACAT	TGACTCCCGC	TTTACGGCAA	840
ATTTAGGTGC	GGGTCAGCGT	TTTTTCCTTC	CTGCAAACAT	GTTGGGCTAT	AACGTCTTCA	900
TTGATCAGGA	TTTTTCTGGT	GATAATACCC	GTTTAGGTAT	TGGTGGCGAA	TACTGGCGAG	960
ACTATTTCAA	AAGTAGCGTT	AACGGCTATT	TCCGCATGAG	GCGCTGGCAT	GAGTCATACC	1020
ATAAGAAAGA	CTATGATGAG	CGCCCAGCAA	ATGGCTTCGA	TATCCGTTTT	AATGGCTATC	1080
TACCGTCATA	TCCGGCATT	GGCGCCAAGC	TGATATATGA	GCAGTATTAT	GGTGATAATG	1140
TTGCTTTGTT	TAATTCTGAT	AAGCTGCAGT	CGAATCCTGG	TGCGGCGACC	GTTGGTGTAA	1200
ACTATACTCC	GATTCCCTCTG	GTGACGATGG	GGATCGATTA	CCGTCATGGT	ACGGGTAATG	1260
AAAATGATCT	CCTTTACTCA	ATGCAGTTCC	GTTATCAGTT	TGATAAATCG	TGGTCTCAGC	1320
AAATTGAACC	ACAGTATGTT	AACGAGTTAA	GAACATTATC	AGGCAGCCGT	TACGATCTGG	1380
TTCAGCGTAA	TAACAATATT	ATTCTGGAGT	ACAAGAAGCA	GGATATTCTT	TCTCTGAATA	1440
TTCCGCATGA	TATTAATGGT	ACTGAACACA	GTACGCAGAA	GATTCAGTTG	ATCGTTAAGA	1500
GCAAATACGG	TCTGGATCGT	ATCGTCTGGG	ATGATAGTGC	ATTACGCAGT	CAGGGCGGTC	1560
AGATTTCAGCA	TAGCGGAAGC	CAAAGCGCAC	AAGACTACCA	GGCTATTTTG	CCTGCTTATG	1620
TGCAAGGTGG	CAGCAATATT	TATAAAGTGA	CGGCTCGCGC	CTATGACCGT	AATGGCAATA	1680
GCTCTAACAA	TGTACAGCTT	ACTATTACCG	TTCTGTCGAA	TGGTCAAGTT	GTCGACCAGG	1740
TTGGGGTAAC	GGACTTTACG	GCGGATAAGA	CTTCGGCTAA	AGCGGATAAC	GCCGATACCA	1800
TTACTTATAC	CGCGACGGTG	AAAAAGAATG	GGGTAGCTCA	GGCTAATGTC	CCTGTTTCAT	1860
TTAATATTGT	TTCAGGAAC	GCAACTCTTG	GGGCAAATAG	TGCCAAAACG	GATGCTAACG	1920
GTAAGGCAAC	CGTAACGTTG	AAGTCGAGTA	CGCCAGGACA	GGTCGTCGTG	TCTGCTAAAA	1980
CCGCGGAGAT	GAGTTCAGCA	CTTAATGCCA	GTGCGGTTAT	ATTTTTTGAT	CAAACCAAGG	2040
CCAGCATTAC	TGAGATTAAG	GCTGATAAGA	CAACTGCAGT	AGCAAATGGT	AAGGATGCTA	2100
TTAAATATAC	TGTAAAAGTT	ATGAAAAACG	GTCAGCCAGT	TAATAATCAA	TCCGTTACAT	2160
TCTCAACAAA	CTTTGGGATG	TTCAACGGTA	AGTCTCAAAC	GCAAGCAACC	ACGGGAAATG	2220
ATGGTCGTGC	GACGATAACA	CTAACTTCCA	GTTCCGCCGG	TAAAGCGACT	GTTAGTGCGA	2280
CAGTCAGTGA	TGGGGCTGAG	GTTAAAGCGA	CTGAGGTCAC	TTTTTTTGAT	GAAGTAAAAA	2340
TTGACAACAA	GGTTGATATT	ATTGGTAACA	ATGTCAGAGG	CGAGTTGCCT	AATATTTGGC	2400

TGCAATATGG TCAGTTTAAA CTGAAAGCAA GCGGTGGTGA TGGTACATAT TCATGGTATT	2460
CAGAAAATAC CAGTATCGCG ACTGTCGATG CATCAGGGAA AGTCACTTTG AATGGTAAAG	2520
GCAGTGTCTG AATTAAAGCC ACATCTGGTG ATAAGCAAAC AGTAAGTTAC ACTATAAAAG	2580
CACCGTCGTA TATGATAAAA GTGGATAAGC AAGCCTATTA TGCTGATGCT ATGTCCATTT	2640
GCAAAAATTT ATTACCATCC ACACAGACGG TATTGTCAGA TATTTATGAC TCATGGGGGG	2700
CTGCAAATAA ATATAGCCAT TATAGTTCTA TGAAGTCAAT AACTGCTTGG ATTAAACAGA	2760
CATCTAGTGA GCAGCGTTCT GGAGTATCAA GCACTTATAA CCTAATAACA CAAAACCCCTC	2820
TTCTGCGGGT TAATGTTAAT ACTCCAAATG TCTATGCGGT TTGTGTAGAA TAATTCCATA	2880
ACCACCCCGG CTAAAATATG TATTGTTTTA GTCGGGGCAT AATTATTTCT TCTTAAGAAA	2940
TAACCTCTTA TAATCAAATC TACTACTGGT CTTTTTATCT GCTTAATAGG TCTCTTTCAA	3000
AGAGACACAT TCACGTTTTT TAGAGTAGGT TGATCCAACC ACGCTGTATA CCAAAGCTGA	3060
ATCACATCAA GCAACAATA TGCTCACAAC ATCCACACAA TAAAAA	3106

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 88 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATGAGAGGAT CGCAYCAYCA YCAYCAYCAY GGATCCGCAT GCGACTCGGT ACCCCGGGTC	60
GACCTGCAGC CAAGCTTAAT TAGCTGAG	88

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 91 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

ATGAGAGGAT CTCAYCAYCA YCAYCAYCAY ACGGATCCGC ATGCGAGCTC GGTACCCCGG	60
GTCGACCTGC AGCCAAGCTT AATTAGCTGA G	91

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 90 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ATGAGAGGAT CTCAYCAYCA YCAYCAYCAY GGGATCCGCA TCGAGAGCTCG GTACCCCGGG	60
TCGACCTGCA GCCAAGCTTA ATTAGCTGAG	90

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 250 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA	60
ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA	120
GGATCGCATC ACCATCACCA TCACGGATCC GCATGCGAGC TCGGTACCCC GGGTCGACCT	180
GCAGCCAAGC TTAATTAGCT GAGCTTGGAC TCCTGTTGAT AGATCCAGTA ATGACCTCAG	240
AACTCCATCT	250